A Pessimistic Approximation for the Fisher Information Measure

Manuel Stein and Josef A. Nossek

***Abstract*—The problem how to determine the intrinsic quality of a signal processing system with respect to the inference of an unknown deterministic parameter θ is considered. While Fisher's information measure $F(\theta)$ forms a classical analytical tool for such a problem, direct computation of the information measure can become difficult in certain situations. This in particular forms an obstacle for the estimation theoretic performance analysis of non-linear measurement systems, where the form of the conditional output probability function can make calculation of the information measure $F(\theta)$ difficult. Based on the Cauchy-Schwarz inequality, we establish an alternative information measure $S(\theta)$. It forms a pessimistic approximation to the Fisher information $F(\theta)$ and has the property that it can be evaluated with the first four output moments at hand. These entities usually exhibit good mathematical tractability or can be determined at low-complexity by output measurements in a calibrated setup or via numerical simulations. With various examples we show that $S(\theta)$ provides a good conservative approximation for $F(\theta)$ and outline different estimation theoretic problems where the presented information bound turns out to be useful.**

***Index Terms*—estimation theory, non-linear systems, Cramér-Rao bound, experimental design, minimum Fisher information, worst-case noise, squaring loss, hard-limiter, soft-limiter.**

I. Introduction

Suppose we are given a parametric system, characterized by a probability density or mass function $q(y;\theta)$, and face the problem of inferring the deterministic but unknown system parameter $\theta \in \Theta$ from measurements at the system output Y. The output Y takes random values $y \in \mathcal{Y}$, where $\mathcal{Y}$ denotes the support of the random variable Y. Estimation theory [1], [2] provides a variety of tools for this kind of problem: On the one hand, guidelines for the design of high-performance *processing algorithms* and on the other hand corresponding *performance bounds* [4]–[9]. While the latter have originally been derived in order to benchmark different estimation algorithms, establish efficiency or identify potential for further improvement, these error bounds have become popular as a figure of merit for the design and optimization of the measurement system $q(y;\theta)$. Such a problem arises frequently in the field of signal processing, where not only the efficient extraction of information from noisy data is within the interest of the engineer, but also the design of the physical measurement system $q(y;\theta)$ itself. Note that the layout of the measurement system can significantly influence technical properties like computational complexity, power consumption, production cost, reliability, processing delay and system performance. Therefore, given the ability to modify the data gathering system $q(y;\theta)$ to an alternative design $p(z;\theta)$ with the altered output Z, exhibiting realizations $z \in \mathcal{Z}$, a rigorous method is required in order to draw a precise conclusion about the intrinsic quality of the original system $q(y;\theta)$ and the envisioned modification $p(z;\theta)$ with respect to the problem of deriving a high performance estimation procedure $\hat{\theta}(\boldsymbol{y})$ or $\hat{\theta}(\boldsymbol{z})$. Here $\boldsymbol{y} \in \mathcal{Y}^N$ and $\boldsymbol{z} \in \mathcal{Z}^N$ denote a collection of N independent realizations of the system outputs Y or Z.

The authors are with the Institute for Circuit Theory and Signal Processing, Technische Universität München, 80290 Germany (e-mail: manuel.stein@tum.de, josef.a.nossek@tum.de).

A. Estimation and Information Measures

We restrict the discussion to unbiased estimation algorithms

$$\int \hat{\theta}(\boldsymbol{y}) q(\boldsymbol{y};\theta) \mathrm{d}\boldsymbol{y} = \theta \tag{1}$$

and assume that the system $q(\boldsymbol{y};\theta)$ is differentiable in $\theta \in \Theta$ for every $\boldsymbol{y} \in \mathcal{Y}^N$, where the parameter set Θ is an open subset on the real line $\mathbb{R}$. Further all considered systems exhibit regularity, such that the statement

$$\frac{\partial}{\partial\theta} \int f(\boldsymbol{y}) q(\boldsymbol{y};\theta) \mathrm{d}\boldsymbol{y} = \int f(\boldsymbol{y}) \frac{\partial q(\boldsymbol{y};\theta)}{\partial\theta} \mathrm{d}\boldsymbol{y} \tag{2}$$

holds for any function $f(\cdot)$ which does not present θ as an argument. Using (1) and (2) we can set out that

$$\int \hat{\theta}(\boldsymbol{y}) \frac{\partial q(\boldsymbol{y};\theta)}{\partial\theta} \mathrm{d}\boldsymbol{y} = 1. \tag{3}$$

With the requirement

$$\int q(\boldsymbol{y};\theta) \mathrm{d}\boldsymbol{y} = 1, \qquad \forall\theta \in \Theta, \tag{4}$$

it follows that

$$\frac{\partial}{\partial\theta} \int q(\boldsymbol{y};\theta) \mathrm{d}\boldsymbol{y} = 0, \qquad \forall\theta \in \Theta, \tag{5}$$

such that we expand (3) by

$$\int (\hat{\theta}(\boldsymbol{y}) - \theta) \frac{\partial q(\boldsymbol{y};\theta)}{\partial\theta} \mathrm{d}\boldsymbol{y} = 1. \tag{6}$$

Using the fact that

$$\frac{\partial \ln q(\boldsymbol{y};\theta)}{\partial\theta} = \frac{1}{q(\boldsymbol{y};\theta)} \frac{\partial q(\boldsymbol{y};\theta)}{\partial\theta}, \tag{7}$$

equation (6) is manipulated, resulting in

$$\int (\hat{\theta}(\boldsymbol{y}) - \theta) \frac{\partial \ln q(\boldsymbol{y};\theta)}{\partial\theta} q(\boldsymbol{y};\theta) \mathrm{d}\boldsymbol{y} = 1. \tag{8}$$

For two real-valued functions $f(\cdot)$ and $g(\cdot)$ the Cauchy-Schwarz inequality [10] states

$$\int f^2(\boldsymbol{x}) p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \int g^2(\boldsymbol{x}) p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \geq \left(\int f(\boldsymbol{x}) g(\boldsymbol{x}) p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right)^2, \tag{9}$$

where equality holds only if

$$f(\boldsymbol{x}) = \kappa g(\boldsymbol{x}) + \lambda, \qquad \forall \boldsymbol{x} \in \mathcal{X}^N \tag{10}$$

with constant $\kappa, \lambda \in \mathbb{R}$. This allows to derive the inequality

$$\int (\hat{\theta}(\boldsymbol{y}) - \theta)^2 q(\boldsymbol{y};\theta) \mathrm{d}\boldsymbol{y} \geq \left(\int \left(\frac{\partial \ln q(\boldsymbol{y};\theta)}{\partial \theta} \right)^2 q(\boldsymbol{y};\theta) \mathrm{d}\boldsymbol{y} \right)^{-1} \tag{11}$$

from expression (8). As long as the observations are independent and identically distributed, i.e., as long as it is possible to factorize

$$q(\boldsymbol{y};\theta) = \prod_{n=1}^{N} q(y_n;\theta), \qquad \forall \boldsymbol{y} \in \mathcal{Y}^N, \tag{12}$$

where y_n denotes the n-th entry in the collection of samples $\boldsymbol{y}$, and each element Y_n follows the identical statistical model

$$q(y_n;\theta) = q(y;\theta), \qquad \forall n \in \{1, 2, \ldots, N\}, \tag{13}$$

the right hand side of (11) simplifies to

$$\begin{aligned} \int \left(\frac{\partial \ln q(\boldsymbol{y};\theta)}{\partial \theta} \right)^2 q(\boldsymbol{y};\theta) \mathrm{d}\boldsymbol{y} &= \\ = N \int_{\mathcal{Y}} \left(\frac{\partial \ln q(y;\theta)}{\partial \theta} \right)^2 q(y;\theta) \mathrm{d}y. \end{aligned} \tag{14}$$

The left hand side of (11) is identified as the mean squared-error $\mathrm{mse}_Y(\theta)$ of the estimator $\hat{\theta}(\boldsymbol{y})$, such that the Cramér-Rao inequality [4] [5] for unbiased estimators

$$\begin{aligned} \mathrm{mse}_Y(\theta) &= \mathrm{var}_Y(\theta) \\ &\geq \frac{1}{N F_Y(\theta)} \end{aligned} \tag{15}$$

is obtained. Consequently, the Fisher information, defined by

$$F_Y(\theta) = \int_{\mathcal{Y}} \left(\frac{\partial \ln q(y;\theta)}{\partial \theta} \right)^2 q(y;\theta) \mathrm{d}y, \tag{16}$$

is a measure for the amount of intrinsic information about the unknown deterministic parameter θ contained in average within each observation of the random output Y. It can be interpreted as the average contribution of each measurement y to the reduction of the uncertainty $\mathrm{var}_Y(\theta)$ about the parameter θ [11]. Note, that the Fisher information measure also plays an important role for performance bounds in the Bayesian setting [12]–[15], where θ is considered to be a random variable. A comprehensive overview on this topic, which is out of the scope of this article, can be found in [16].

B. Relative Inference Capability

As the inequality (15) holds for all estimation procedures satisfying (1) and asymptotically in N attains equality when the estimator $\hat{\theta}(\boldsymbol{y})$ is efficient, the Fisher information measure (16) can be used to unambiguously assess the relative estimation theoretic quality of the modification $p(z;\theta)$ with respect to the reference $q(y;\theta)$ by the information ratio

$$\chi(\theta) = \frac{F_Z(\theta)}{F_Y(\theta)}. \tag{17}$$

Note that $F_Z(\theta)$ is the Fisher information (16) evaluated on $\mathcal{Z}$ with respect to the conditional probability function $p(z;\theta)$.

C. Fisher Information Bound

Using $\chi(\theta)$ for the design and optimization of the measurement system requires to compute (16) for the benchmark experiment $q(y;\theta)$ and all modifications $p(z;\theta)$ which are of interest. If $p(z;\theta)$ takes a complicated form this can become difficult. In a situation where the parametric model $p(z;\theta)$ governing the statistics of the output Z is unknown, a direct analytical formulation of the information measure (16) becomes impossible. However, if the first moment

$$\mu_1(\theta) = \int_{\mathcal{Z}} z p_z(z;\theta) \mathrm{d}z \tag{18}$$

of the system output Z and the second central output moment

$$\mu_2(\theta) = \int_{\mathcal{Z}} \big(z - \mu_1(\theta)\big)^2 p_z(z;\theta) \mathrm{d}z, \tag{19}$$

are known and are both differentiable in θ, it has recently been shown that the Fisher information $F(\theta)$ is in general bounded from below [17]

$$F_Z(\theta) \geq \frac{1}{\mu_2(\theta)} \left(\frac{\partial \mu_1(\theta)}{\partial \theta} \right)^2. \tag{20}$$

While examples can be given where (20) holds with equality [17], a simple counter example is immediately constructed. To this end, consider the system output to follow the generic parametric Gaussian distribution

$$p(z;\theta) = \frac{1}{\sqrt{2\pi\mu_2(\theta)}} \mathrm{e}^{-\frac{(z-\mu_1(\theta))^2}{2\mu_2(\theta)}}. \tag{21}$$

The exact Fisher information is [3, pp. 47]

$$F_Z(\theta) = \frac{1}{\mu_2(\theta)} \left(\frac{\partial \mu_1(\theta)}{\partial \theta} \right)^2 + \frac{1}{2\mu_2^2(\theta)} \left(\frac{\partial \mu_2(\theta)}{\partial \theta} \right)^2, \tag{22}$$

and is equal to (20) only for the special case where

$$\frac{\partial \mu_2(\theta)}{\partial \theta} = 0. \tag{23}$$

Obviously the inequality (20) does in general not take into account the contribution provided by the variation of the second output moment $\mu_2(\theta)$ to the Fisher information measure $F_Z(\theta)$.

D. Contribution and Outline

Motivated by this insight, we aim at a substantial improvement of our lower bound for $F(\theta)$, which we provided in our previous discussion [17]. We achieve this by utilizing the Cauchy-Schwarz inequality (9) under a generalized approach and subsequently maximizing the resulting expression in order to attain an alternative information measure $S(\theta)$. The proposed pessimistic approximation for $F(\theta)$ exclusively contains the first four output moments in parametric form. A discussion for situations like (23) shows that the inequality (20) is contained in the result as one special case. Using various examples with continuous and discrete system outputs, we verify the quality of the alternative information measure $S(\theta)$. In order to demonstrate possible applications of the result and further insights, through $S(\theta)$ we approximately determine the estimation theoretic information loss when squaring a standard

Gaussian input distribution and advance on the discussion about minimum Fisher information [11], [17]–[21]. Finally, we mimic a situation of practical relevance. Measuring the output moments of a soft-limiting device with standard Gaussian input, we demonstrate how to conservatively establish the intrinsic inference capability $F(\theta)$ of a non-linear signal processing system when the analytic form of the parametric output statistic $p(z;\theta)$ is not directly available.

II. Improved Fisher Information Bound

For the discussion we additionally require the central output moments

$$\begin{aligned} \mu_3(\theta) &= \int_{\mathcal{Z}} \big(z-\mu_1(\theta)\big)^3 p(z;\theta)\mathrm{d}z \\ \mu_4(\theta) &= \int_{\mathcal{Z}} \big(z-\mu_1(\theta)\big)^4 p(z;\theta)\mathrm{d}z \end{aligned} \tag{24}$$

and their normalized versions

$$\begin{aligned} \bar{\mu}_3(\theta) &= \int_{\mathcal{Z}} \left(\frac{z-\mu_1(\theta)}{\sqrt{\mu_2(\theta)}}\right)^3 p_z(z;\theta)\mathrm{d}z \\ &= \mu_3(\theta)\mu_2^{-\frac{3}{2}}(\theta) \end{aligned} \tag{25}$$

$$\begin{aligned} \bar{\mu}_4(\theta) &= \int_{\mathcal{Z}} \left(\frac{z-\mu_1(\theta)}{\sqrt{\mu_2(\theta)}}\right)^4 p_z(z;\theta)\mathrm{d}z \\ &= \mu_4(\theta)\mu_2^{-2}(\theta). \end{aligned} \tag{26}$$

Note that $\bar{\mu}_3(\theta)$ is refereed to as the skewness, an indicator for the asymmetry of the output distribution $p(z;\theta)$, while $\bar{\mu}_4(\theta)$ is called the kurtosis, a characterization for the shape of the output distribution $p(z;\theta)$. Both moments stand in relation through Pearson's inequality [22]

$$\bar{\mu}_4(\theta) \geq \bar{\mu}_3^2(\theta) + 1. \tag{27}$$

A compact and elegant proof on (27) can be found in [23].

A. Generalized Bounding Approach

We apply the inequality (9) with

$$f(z;\theta) = \frac{\partial \ln p(z;\theta)}{\partial \theta} \tag{28}$$

and

$$g(z;\theta) = \left(\frac{z-\mu_1(\theta)}{\sqrt{\mu_2(\theta)}}\right) + \beta(\theta)\left(\frac{z-\mu_1(\theta)}{\sqrt{\mu_2(\theta)}}\right)^2 - \beta(\theta), \tag{29}$$

where $\beta(\theta) \in \mathbb{R}$, in order to lower bound the Fisher information

$$F(\theta) = \int_{\mathcal{Z}} f^2(z;\theta)p(z;\theta)\mathrm{d}z. \tag{30}$$

With the manipulations

$$\begin{aligned} &\int_{\mathcal{Z}} \left(\frac{z-\mu_1(\theta)}{\sqrt{\mu_2(\theta)}}\right) \frac{\partial \ln p_z(z;\theta)}{\partial \theta} p_z(z;\theta)\mathrm{d}z = \\ &= \frac{1}{\sqrt{\mu_2(\theta)}}\left(\int_{\mathcal{Z}} z\frac{\partial p_z(z;\theta)}{\partial \theta}\mathrm{d}z - \mu_1(\theta)\int_{\mathcal{Z}} \frac{\partial p_z(z;\theta)}{\partial \theta}\mathrm{d}z\right) \\ &= \frac{1}{\sqrt{\mu_2(\theta)}}\left(\frac{\partial}{\partial \theta}\int_{\mathcal{Z}} z p_z(z;\theta)\mathrm{d}z - \mu_1(\theta)\frac{\partial}{\partial \theta}\int_{\mathcal{Z}} p_z(z;\theta)\mathrm{d}z\right) \\ &= \frac{1}{\sqrt{\mu_2(\theta)}}\frac{\partial \mu_1(\theta)}{\partial \theta} \end{aligned} \tag{31}$$

and

$$\begin{aligned} &\int_{\mathcal{Z}} \left(\frac{z-\mu_1(\theta)}{\sqrt{\mu_2(\theta)}}\right)^2 \frac{\partial \ln p_z(z;\theta)}{\partial \theta} p_z(z;\theta)\mathrm{d}z = \\ &= \frac{1}{\mu_2(\theta)}\Bigg(\int_{\mathcal{Z}} z^2\frac{\partial p_z(z;\theta)}{\partial \theta}\mathrm{d}z - 2\mu_1(\theta)\int_{\mathcal{Z}} z\frac{\partial p_z(z;\theta)}{\partial \theta}\mathrm{d}z \\ &\qquad + \mu_1^2(\theta)\int_{\mathcal{Z}} \frac{\partial p_z(z;\theta)}{\partial \theta}\mathrm{d}z\Bigg) \\ &= \frac{1}{\mu_2(\theta)}\left(\frac{\partial}{\partial \theta}\int_{\mathcal{Z}} z^2 p_z(z;\theta)\mathrm{d}z - 2\mu_1(\theta)\frac{\partial}{\partial \theta}\int_{\mathcal{Z}} z p_z(z;\theta)\mathrm{d}z\right) \\ &= \frac{1}{\mu_2(\theta)}\left(\frac{\partial}{\partial \theta}\big(\mu_2(\theta)+\mu_1^2(\theta)\big) - 2\mu_1(\theta)\frac{\partial \mu_1(\theta)}{\partial \theta}\right) \\ &= \frac{1}{\mu_2(\theta)}\frac{\partial \mu_2(\theta)}{\partial \theta}, \end{aligned} \tag{32}$$

where we use the fact that

$$\int_{\mathcal{Z}} z^2 p_z(z;\theta)\mathrm{d}z = \mu_2(\theta) + \mu_1^2(\theta), \tag{33}$$

the identity

$$\begin{aligned} &\int_{\mathcal{Z}} f(z;\theta)g(z;\theta)p(z;\theta)\mathrm{d}z = \\ &\qquad = \frac{1}{\sqrt{\mu_2(\theta)}}\frac{\partial \mu_1(\theta)}{\partial \theta} + \frac{\beta(\theta)}{\mu_2(\theta)}\frac{\partial \mu_2(\theta)}{\partial \theta}, \end{aligned} \tag{34}$$

is found. Note that

$$\begin{aligned} \int_{\mathcal{Z}} \beta(\theta)\frac{\partial \ln p(z;\theta)}{\partial \theta}p(z;\theta)\mathrm{d}z &= \\ &= \beta(\theta)\int_{\mathcal{Z}} \frac{\partial \ln p(z;\theta)}{\partial \theta}p(z;\theta)\mathrm{d}z \\ &= \beta(\theta)\frac{\partial}{\partial \theta}\int_{\mathcal{Z}} p(z;\theta)\mathrm{d}z \\ &= 0. \end{aligned} \tag{35}$$

Taking into account that

$$\int_{\mathcal{Z}} \left(\frac{z-\mu_1(\theta)}{\sqrt{\mu_2(\theta)}}\right)p(z;\theta)\mathrm{d}z = 0, \tag{36}$$

we get

$$\begin{aligned} &\int_{\mathcal{Z}} g^2(z;\theta)p(z;\theta)\mathrm{d}z = \\ &\qquad = 1 + 2\beta(\theta)\bar{\mu}_3(\theta) + \beta^2(\theta)\bar{\mu}_4(\theta) - \beta^2(\theta). \end{aligned} \tag{37}$$

Therefore, from (9), (30), (34) and (37) it can be shown, that the Fisher information can in general not fall below

$$F(\theta) \geq \frac{\left(\int_{\mathcal{Z}} f(z;\theta) g(z;\theta) p(z;\theta) \mathrm{d}z\right)^2}{\int_{\mathcal{Z}} g^2(z;\theta) p(z;\theta) \mathrm{d}z} = \frac{1}{\mu_2(\theta)} \frac{\left(\frac{\partial \mu_1(\theta)}{\partial \theta} + \frac{\beta}{\sqrt{\mu_2(\theta)}} \frac{\partial \mu_2(\theta)}{\partial \theta}\right)^2}{1 + 2\beta(\theta)\bar{\mu}_3(\theta) + \beta^2(\theta)(\bar{\mu}_4(\theta) - 1)}. \quad (38)$$

B. Optimization of the Information Bound

The expression (38) contains the factor $\beta(\theta)$ which can be used to improve the lower bound. For the trivial choice of $\beta(\theta) = 0$, the expression becomes

$$F(\theta) \geq \frac{1}{\mu_2(\theta)} \left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2, \quad (39)$$

which turns out to be the bound (20) discussed in [17]. In order to improve this result, consider that the problem

$$x^\star = \arg\max_{x \in \mathbb{R}} h(x) \quad (40)$$

with

$$h(x) = \frac{(a + xb)^2}{1 + 2xc + x^2 d}, \quad (41)$$

has a unique maximizing solution

$$x^\star = \frac{ac - b}{bc - ad}. \quad (42)$$

Consequently, the tightest form of (38) is given by

$$F(\theta) \geq \frac{1}{\mu_2(\theta)} \frac{\left(\frac{\partial \mu_1(\theta)}{\partial \theta} + \frac{\beta^\star(\theta)}{\sqrt{\mu_2(\theta)}} \frac{\partial \mu_2(\theta)}{\partial \theta}\right)^2}{1 + 2\beta^\star(\theta)\bar{\mu}_3(\theta) + \beta^{\star 2}(\theta)(\bar{\mu}_4(\theta) - 1)} = S(\theta), \quad (43)$$

with the optimization factor

$$\beta^\star(\theta) = \frac{\frac{\partial \mu_1(\theta)}{\partial \theta}\bar{\mu}_3(\theta) - \frac{1}{\sqrt{\mu_2(\theta)}}\frac{\partial \mu_2(\theta)}{\partial \theta}}{\frac{1}{\sqrt{\mu_2(\theta)}}\frac{\partial \mu_2(\theta)}{\partial \theta}\bar{\mu}_3(\theta) - \frac{\partial \mu_1(\theta)}{\partial \theta}(\bar{\mu}_4(\theta) - 1)} = \frac{\frac{\partial \mu_1(\theta)}{\partial \theta}\sqrt{\mu_2(\theta)}\bar{\mu}_3(\theta) - \frac{\partial \mu_2(\theta)}{\partial \theta}}{\frac{\partial \mu_2(\theta)}{\partial \theta}\bar{\mu}_3(\theta) - \frac{\partial \mu_1(\theta)}{\partial \theta}\sqrt{\mu_2(\theta)}(\bar{\mu}_4(\theta) - 1)}. \quad (44)$$

The inequality (43) states that the derived information measure $S(\theta)$ is always dominated by the Fisher information measure $F(\theta)$. Therefore, $S(\theta)$ gives a cautious approximation for $F(\theta)$. Note that the Fisher information $F(\theta)$ requires to integrate the squared-score $\left(\frac{\partial \ln p(z;\theta)}{\partial \theta}\right)^2$. In contrast, the alternative measure $S(\theta)$ exclusively requires the first four central output moments $\mu_1(\theta)$, $\mu_2(\theta)$, $\bar{\mu}_3(\theta)$, $\bar{\mu}_4(\theta)$ in parametric form.

III. Fisher Information Bound - Special Cases

In order to derive simplified forms of the presented information measure $S(\theta)$, let us consider some special cases.

A. Constant First Moment

For the situation where the first moment $\mu_1(\theta)$ does not vary with the system parameter θ, i.e.,

$$\frac{\partial \mu_1(\theta)}{\partial \theta} = 0, \qquad \forall \theta \in \Theta, \quad (45)$$

we attain

$$\beta^\star(\theta) = -\frac{1}{\bar{\mu}_3(\theta)}, \quad (46)$$

such that a pessimistic approximation for $F(\theta)$ is

$$S(\theta) = \frac{1}{\mu_2(\theta)} \frac{\left(-\frac{1}{\bar{\mu}_3(\theta)\sqrt{\mu_2(\theta)}}\frac{\partial \mu_2(\theta)}{\partial \theta}\right)^2}{1 - 2 + \frac{(\bar{\mu}_4(\theta) - 1)}{\bar{\mu}_3^2(\theta)}} = \frac{1}{\mu_2^2(\theta)} \frac{\left(\frac{\partial \mu_2(\theta)}{\partial \theta}\right)^2}{\bar{\mu}_4(\theta) - \bar{\mu}_3^2(\theta) - 1}. \quad (47)$$

Note that inequality (27) assures that $S(\theta)$ stays positive under these circumstances.

B. Constant Second Moment

When the second moment $\mu_2(\theta)$ is constant within θ, i.e.,

$$\frac{\partial \mu_2(\theta)}{\partial \theta} = 0, \qquad \forall \theta \in \Theta, \quad (48)$$

it holds that

$$\beta^\star(\theta) = -\frac{\bar{\mu}_3(\theta)}{(\bar{\mu}_4(\theta) - 1)}. \quad (49)$$

In this situation

$$S(\theta) = \frac{1}{\mu_2(\theta)} \frac{\left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2}{1 - 2\frac{\bar{\mu}_3^2(\theta)}{(\bar{\mu}_4(\theta) - 1)} + \frac{\bar{\mu}_3^2(\theta)}{(\bar{\mu}_4(\theta) - 1)}} = \frac{1}{\mu_2(\theta)} \frac{\left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2}{1 - \frac{\bar{\mu}_3^2(\theta)}{(\bar{\mu}_4(\theta) - 1)}}. \quad (50)$$

Note that (50) equals the expression in (20) whenever the skewness $\bar{\mu}_3$ vanishes. In general the relation (27) makes (50) larger than the unoptimized bound (20).

C. Symmetric Distributions

For symmetric output distributions with zero skewness, i.e.,

$$\bar{\mu}_3(\theta) = 0, \quad (51)$$

we verify that the optimization of the information bound derived in (43) results in

$$\beta^\star(\theta) = \frac{\frac{\partial \mu_2(\theta)}{\partial \theta}}{\frac{\partial \mu_1(\theta)}{\partial \theta}\sqrt{\mu_2(\theta)}(\bar{\mu}_4(\theta) - 1)}, \quad (52)$$

such that

$$S(\theta) = \frac{1}{\mu_2(\theta)} \frac{\left(\frac{\partial \mu_1(\theta)}{\partial \theta} + \frac{\left(\frac{\partial \mu_2(\theta)}{\partial \theta}\right)^2}{\frac{\partial \mu_1(\theta)}{\partial \theta}\mu_2(\theta)(\bar{\mu}_4(\theta)-1)}\right)^2}{1 + \left(\frac{\frac{\partial \mu_2(\theta)}{\partial \theta}}{\frac{\partial \mu_1(\theta)}{\partial \theta}\sqrt{\mu_2(\theta)(\bar{\mu}_4(\theta)-1)}}\right)^2(\bar{\mu}_4(\theta)-1)}$$
$$= \frac{\left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2 \mu_2(\theta)(\bar{\mu}_4(\theta)-1) + \left(\frac{\partial \mu_2(\theta)}{\partial \theta}\right)^2}{\mu_2^2(\theta)(\bar{\mu}_4(\theta)-1)}$$
$$= \frac{1}{\mu_2(\theta)}\left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2 + \frac{1}{\mu_2^2(\theta)(\bar{\mu}_4(\theta)-1)}\left(\frac{\partial \mu_2(\theta)}{\partial \theta}\right)^2. \tag{53}$$

Again note that according to Pearson's inequality (27)

$$\bar{\mu}_4(\theta) - 1 \geq 0, \tag{54}$$

such that the expression (53) always takes a positive value.

D. Simplifying Characteristic

For the case where the identity

$$\frac{\partial \mu_1(\theta)}{\partial \theta}\sqrt{\mu_2(\theta)}\bar{\mu}_3(\theta) = \frac{\partial \mu_2(\theta)}{\partial \theta}, \qquad \forall \theta \in \Theta, \tag{55}$$

holds, the optimization of (43) results in

$$\beta^\star(\theta) = 0 \tag{56}$$

and the approximation obtains the compact form (20)

$$S(\theta) = \frac{1}{\mu_2(\theta)}\left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2. \tag{57}$$

This situation occurs for example for a symmetric output distribution with constant second moment.

IV. Approximation Quality - Continuous Outputs

In order to demonstrate the quality of the derived lower bound $S(\theta)$, we consider different examples where $F(\theta)$ can be derived in compact form. First we discuss several well-studied distributions with continuous support $\mathcal{Z}$.

A. Gaussian System Output

Consider the system output Z to be the undisturbed observation of a generic Gaussian distribution in parametric form

$$p(z;\theta) = \frac{1}{\sqrt{2\pi\nu_2(\theta)}} \mathrm{e}^{-\frac{(z-\nu_1(\theta))^2}{2\nu_2(\theta)}}. \tag{58}$$

The exact Fisher information measure is given by

$$F(\theta) = \frac{1}{\nu_2(\theta)}\left(\frac{\partial \nu_1(\theta)}{\partial \theta}\right)^2 + \frac{1}{2\nu_2^2(\theta)}\left(\frac{\partial \nu_2(\theta)}{\partial \theta}\right)^2. \tag{59}$$

As for this case the output moments of interest are

$$\begin{aligned} \mu_1(\theta) &= \nu_1(\theta) \\ \mu_2(\theta) &= \nu_2(\theta) \\ \bar{\mu}_3(\theta) &= 0 \\ \bar{\mu}_4(\theta) &= 3, \end{aligned} \tag{60}$$

we get the approximation

$$S(\theta) = \frac{1}{\mu_2(\theta)}\left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2 + \frac{1}{\mu_2^2(\theta)(\bar{\mu}_4(\theta)-1)}\left(\frac{\partial \mu_2(\theta)}{\partial \theta}\right)^2$$
$$= \frac{1}{\nu_2(\theta)}\left(\frac{\partial \nu_1(\theta)}{\partial \theta}\right)^2 + \frac{1}{2\nu_2^2(\theta)}\left(\frac{\partial \nu_2(\theta)}{\partial \theta}\right)^2, \tag{61}$$

which is obviously a tight lower bound for the original information measure $F(\theta)$.

B. Exponential System Output

As another example we analyze the case where samples from a parametric exponential distribution

$$p(z;\theta) = \nu(\theta)\mathrm{e}^{-\nu(\theta)z}, \tag{62}$$

with $\nu(\theta) \geq 0$ and $z \geq 0$, can be collected at the random system output Z. The score function under this model is

$$\frac{\partial \ln p(z;\theta)}{\partial \theta} = \frac{1}{\nu(\theta)}\frac{\partial \nu(\theta)}{\partial \theta} - z\frac{\partial \nu(\theta)}{\partial \theta}, \tag{63}$$

such that the Fisher information is evaluated to be

$$F(\theta) = \int_{\mathcal{Z}} \left(\frac{\partial \ln p_z(z;\theta)}{\partial \theta}\right)^2 p_z(z;\theta)\mathrm{d}z$$
$$= \frac{1}{\nu^2(\theta)}\left(\frac{\partial \nu(\theta)}{\partial \theta}\right)^2. \tag{64}$$

For the approximation $S(\theta)$ the required moments are

$$\begin{aligned} \mu_1(\theta) &= \frac{1}{\nu(\theta)} \\ \mu_2(\theta) &= \frac{1}{\nu^2(\theta)} \\ \bar{\mu}_3(\theta) &= 2 \\ \bar{\mu}_4(\theta) &= 3, \end{aligned} \tag{65}$$

such that

$$\frac{\partial \mu_1(\theta)}{\partial \theta}\sqrt{\mu_2(\theta)}\bar{\mu}_3(\theta) = -\frac{2}{\nu^3(\theta)}\frac{\partial \nu(\theta)}{\partial \theta}$$
$$= \frac{\partial \mu_2(\theta)}{\partial \theta}, \tag{66}$$

producing $\beta^\star(\theta) = 0$. The approximation is therefore given by the simplified form

$$S(\theta) = \frac{1}{\mu_2(\theta)}\left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2$$
$$= \nu^2(\theta)\left(-\frac{1}{\nu^2(\theta)}\frac{\partial \nu(\theta)}{\partial \theta}\right)^2$$
$$= \frac{1}{\nu^2(\theta)}\left(\frac{\partial \nu(\theta)}{\partial \theta}\right)^2, \tag{67}$$

which obviously matches the true Fisher information $F(\theta)$ in (64) exactly.

C. Laplacian System Output

For a third example, we assume that the output Z follows a parametric Laplace distribution with zero mean, i.e.,

$$p(z;\theta) = \frac{1}{2\nu(\theta)} \mathrm{e}^{-\frac{|z|}{\nu(\theta)}}. \tag{68}$$

The score function is given by

$$\frac{\partial \ln p(z;\theta)}{\partial \theta} = -\frac{1}{\nu(\theta)} \frac{\partial \nu(\theta)}{\partial \theta} + \frac{|z|}{\nu^2(\theta)} \frac{\partial \nu(\theta)}{\partial \theta} \tag{69}$$

and the exact Fisher information is found to be

$$\begin{aligned} F(\theta) &= \int_{\mathcal{Z}} \left(\frac{\partial \ln p_z(z;\theta)}{\partial \theta} \right)^2 p_z(z;\theta) \mathrm{d}z \\ &= \frac{1}{\nu^2(\theta)} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2. \end{aligned} \tag{70}$$

The first four moments of the output Z are

$$\begin{aligned} \mu_1(\theta) &= 0 \\ \mu_2(\theta) &= 2\nu^2(\theta) \\ \bar{\mu}_3(\theta) &= 0 \\ \bar{\mu}_4(\theta) &= 6. \end{aligned} \tag{71}$$

As the first moment is constant with respect to the system parameter θ, the approximation takes the form

$$\begin{aligned} S(\theta) &= \frac{1}{\mu_2^2(\theta)} \frac{\left(\frac{\partial \mu_2(\theta)}{\partial \theta} \right)^2}{(\bar{\mu}_4(\theta) - 1)} \\ &= \frac{1}{4\nu^4(\theta)} \frac{\left(4\nu(\theta) \frac{\partial \nu(\theta)}{\partial \theta} \right)^2}{5} \\ &= \frac{4}{5} \frac{1}{\nu^2(\theta)} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2. \end{aligned} \tag{72}$$

In contrast to the other examples, the information bound $S(\theta)$ is not tight under the Laplacian system model. However, $S(\theta)$ still allows to obtain a pessimistic characterization for the Fisher information measure $F(\theta)$.

V. Approximation Quality - Discrete Outputs

In the following we extend the discussion on the bounding quality of $S(\theta)$ to the case where the system output Z takes values from a discrete alphabet $\mathcal{Z}$.

A. Bernoulli System Output

As a first example for such kind of system outputs, observations from a parametric Bernoulli distribution with

$$\begin{aligned} p(z=1;\theta) &= 1 - p(z=0;\theta) \\ &= \nu(\theta), \end{aligned} \tag{73}$$

are considered, where

$$0 < \nu(\theta) < 1, \qquad \forall \theta \in \Theta. \tag{74}$$

The Fisher information measure under this model is

$$\begin{aligned} F(\theta) &= \int_{\mathcal{Z}} \left(\frac{\partial \ln p(z;\theta)}{\partial \theta} \right)^2 p(z;\theta) \mathrm{d}z \\ &= \sum_{\mathcal{Z}} \left(\frac{\partial p(z;\theta)}{\partial \theta} \right)^2 \frac{1}{p(z;\theta)} \\ &= \frac{\left(\frac{\partial p(z=1;\theta)}{\partial \theta} \right)^2}{p(z=1;\theta)} + \frac{\left(\frac{\partial p(z=0;\theta)}{\partial \theta} \right)^2}{p(z=0;\theta)} \\ &= \frac{1}{\nu(\theta)(1-\nu(\theta))} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2. \end{aligned} \tag{75}$$

The first two moments are

$$\begin{aligned} \mu_1(\theta) &= \nu(\theta) \\ \mu_2(\theta) &= \nu(\theta)(1-\nu(\theta)), \end{aligned} \tag{76}$$

with derivatives

$$\begin{aligned} \frac{\partial \mu_1(\theta)}{\partial \theta} &= \frac{\partial \nu(\theta)}{\partial \theta} \\ \frac{\partial \mu_2(\theta)}{\partial \theta} &= \big(1 - 2\nu(\theta)\big) \frac{\partial \nu(\theta)}{\partial \theta}. \end{aligned} \tag{77}$$

The third normalized moment is

$$\begin{aligned} \bar{\mu}_3(\theta) &= \sum_{\mathcal{Z}} \left(\frac{z - \mu_1(\theta)}{\sqrt{\mu_2(\theta)}} \right)^3 p(z;\theta) \\ &= \left(\frac{1-\nu(\theta)}{\sqrt{\nu(\theta)(1-\nu(\theta))}} \right)^3 \nu(\theta) \\ &\quad + \left(\frac{-\nu(\theta)}{\sqrt{\nu(\theta)(1-\nu(\theta))}} \right)^3 (1-\nu(\theta)) \\ &= \frac{1-2\nu(\theta)}{\sqrt{\nu(\theta)(1-\nu(\theta))}} \end{aligned} \tag{78}$$

and the fourth normalized moment

$$\begin{aligned} \bar{\mu}_4(\theta) &= \sum_{\mathcal{Z}} \left(\frac{z - \mu_1(\theta)}{\sqrt{\mu_2(\theta)}} \right)^4 p(z;\theta) \\ &= \left(\frac{1-\nu(\theta)}{\sqrt{\nu(\theta)(1-\nu(\theta))}} \right)^4 \nu(\theta) \\ &\quad + \left(\frac{-\nu(\theta)}{\sqrt{\nu(\theta)(1-\nu(\theta))}} \right)^4 (1-\nu(\theta)) \\ &= \frac{1}{\nu(\theta)(1-\nu(\theta))} - 3. \end{aligned} \tag{79}$$

As

$$\begin{aligned} \frac{\partial \mu_1(\theta)}{\partial \theta} \sqrt{\mu_2(\theta)} \bar{\mu}_3(\theta) &= \big(1 - 2\nu(\theta)\big) \frac{\partial \nu(\theta)}{\partial \theta} \\ &= \frac{\partial \mu_2(\theta)}{\partial \theta} \end{aligned} \tag{80}$$

and consequently $\beta^\star(\theta) = 0$, the approximation takes its simplified form

$$\begin{aligned} S(\theta) &= \frac{1}{\mu_2(\theta)} \left(\frac{\partial \mu_1(\theta)}{\partial \theta} \right)^2 \\ &= \frac{1}{\nu(\theta)(1-\nu(\theta))} \left(\frac{\partial \nu(\theta)}{\partial \theta} \right)^2. \end{aligned} \tag{81}$$

It becomes clear that also for a binary system output Z, following a parametric Bernoulli distribution, the derived expression $S(\theta)$ is a tight approximation for the original inference capability $F(\theta)$.

B. Poissonian System Output

As a second example with discrete output, we consider the Poisson distribution. The samples z at the output Z are distributed according to the model

$$p(z;\theta) = \frac{\nu^{z}(\theta)}{z!}\mathrm{e}^{-\nu(\theta)}, \tag{82}$$

with

$$\mathcal{Z} = \{0,1,2,\dots\} \tag{83}$$

and

$$\nu(\theta) > 0, \qquad \forall\theta\in\Theta. \tag{84}$$

The second derivative of the log-likelihood is given by

$$\frac{\partial^2 \ln p_z(z;\theta)}{\partial\theta^2} = z\left(\frac{\frac{\partial^2\nu(\theta)}{\partial\theta^2}\nu(\theta) - \left(\frac{\partial\nu(\theta)}{\partial\theta}\right)^2}{\nu^2(\theta)}\right) - \frac{\partial^2\nu(\theta)}{\partial\theta^2}. \tag{85}$$

With the mean of the system output being

$$\mathrm{E}\left[Z\right] = \nu(\theta), \tag{86}$$

we calculate

$$\begin{aligned} F(\theta) &= \int_{\mathcal{Z}}\left(\frac{\partial\ln p(z;\theta)}{\partial\theta}\right)^2 p(z;\theta)\mathrm{d}z \\ &= -\int_{\mathcal{Z}}\frac{\partial^2\ln p(z;\theta)}{\partial\theta^2}p(z;\theta)\mathrm{d}z \\ &= \frac{1}{\nu(\theta)}\left(\frac{\partial\nu(\theta)}{\partial\theta}\right)^2. \end{aligned} \tag{87}$$

In order to apply the approximation $S(\theta)$, we require the moments which are given by

$$\begin{aligned} \mu_1(\theta) &= \nu(\theta) \\ \mu_2(\theta) &= \nu(\theta) \\ \bar{\mu}_3(\theta) &= \frac{1}{\sqrt{\nu(\theta)}} \\ \bar{\mu}_4(\theta) &= \frac{1}{\nu(\theta)} + 3. \end{aligned} \tag{88}$$

As these quantities exhibit the property

$$\begin{aligned} \frac{\partial\mu_1(\theta)}{\partial\theta}\sqrt{\mu_2(\theta)}\bar{\mu}_3(\theta) &= \frac{\partial\nu(\theta)}{\partial\theta} \\ &= \frac{\partial\mu_2(\theta)}{\partial\theta}, \end{aligned} \tag{89}$$

we obtain $\beta^{\star}(\theta) = 0$ and the approximation for this example

$$\begin{aligned} S(\theta) &= \frac{1}{\mu_2(\theta)}\Big(\frac{\partial\mu_1(\theta)}{\partial\theta}\Big)^2 \\ &= \frac{1}{\nu(\theta)}\Big(\frac{\partial\nu(\theta)}{\partial\theta}\Big)^2 \end{aligned} \tag{90}$$

is tight with respect to $F(\theta)$.

C. Hard-limited Gaussian System Output

As a last discrete example, we consider the output Z of a hard-limiting device [24], i.e.,

$$Z = \operatorname{sign}_{\gamma}(Y), \tag{91}$$

where the generalized signum operator is defined by

$$\operatorname{sign}_{\gamma}(x) = \begin{cases} +1 & \text{if } x \geq \gamma \\ -1 & \text{if } x < \gamma. \end{cases} \tag{92}$$

As input Y to the hard-limiter, a generic parametric Gaussian distribution

$$p(y;\theta) = \frac{1}{\sqrt{2\pi\nu_2(\theta)}}\mathrm{e}^{-\frac{(y-\nu_1(\theta))^2}{2\nu_2(\theta)}} \tag{93}$$

is used. The conditional probability mass function of the binary output Z in this experiment is

$$p(z=1;\theta) = \mathrm{Q}\left(\frac{\gamma-\nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right) \tag{94}$$

$$p(z=-1;\theta) = 1-\mathrm{Q}\left(\frac{\gamma-\nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right) \tag{95}$$

with

$$\mathrm{Q}\left(x\right) = \frac{1}{\sqrt{2\pi}}\int_{x}^{\infty}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t \tag{96}$$

being the Q-function. Note that the derivative of the Q-function is given by

$$\frac{\partial\,\mathrm{Q}\left(x\right)}{\partial x} = -\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}. \tag{97}$$

The corresponding derivatives of the conditional probability mass function in this example take the form

$$\begin{aligned} &\frac{\partial p(z=1;\theta)}{\partial\theta} = \\ &\quad= \frac{\mathrm{e}^{-\frac{(\gamma-\nu_1(\theta))^2}{2\mu_2(\theta)}}\left(\frac{\partial\nu_1(\theta)}{\partial\theta}\sqrt{\nu_2(\theta)} + \frac{(\gamma-\nu_1(\theta))}{2\sqrt{\nu_2(\theta)}}\frac{\partial\nu_2(\theta)}{\partial\theta}\right)}{\sqrt{2\pi}\nu_2(\theta)} \end{aligned} \tag{98}$$

and

$$\frac{\partial p(z=-1;\theta)}{\partial\theta} = -\frac{\partial p(z=1;\theta)}{\partial\theta}. \tag{99}$$

Thus, the exact Fisher information $F(\theta)$ is found to be

$$\begin{aligned} F(\theta) &= \frac{\left(\frac{\partial p(z=1;\theta)}{\partial\theta}\right)^2}{p(z=1;\theta)} + \frac{\left(\frac{\partial p(z=-1;\theta)}{\partial\theta}\right)^2}{p(z=-1;\theta)} \\ &= \frac{\mathrm{e}^{-\frac{(\gamma-\nu_1(\theta))^2}{\mu_2(\theta)}}\left(\frac{\partial\nu_1(\theta)}{\partial\theta}\sqrt{\nu_2(\theta)} + \frac{(\gamma-\nu_1(\theta))}{2\sqrt{\nu_2(\theta)}}\frac{\partial\nu_2(\theta)}{\partial\theta}\right)^2}{2\pi\nu_2^2(\theta)\,\mathrm{Q}\left(\frac{\gamma-\nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)\left(1-\mathrm{Q}\left(\frac{\gamma-\nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)\right)}. \end{aligned} \tag{100}$$

For the approximation $S(\theta)$, we calculate the first two output moments by

$$\begin{aligned} \mu_1(\theta) &= p(z=1;\theta) - p(z=-1;\theta) \\ &= 2\,\mathrm{Q}\left(\frac{\gamma-\nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right) - 1 \end{aligned} \tag{101}$$

and

$$\mu_2(\theta) = \sum_{\mathcal{Z}} (z - \mu_1(\theta))^2 p(z;\theta)$$
$$= 4\left(1 - \mathrm{Q}\left(\frac{\gamma - \nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)\right)^2 \mathrm{Q}\left(\frac{\gamma - \nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)$$
$$+ 4\left(1 - \mathrm{Q}\left(\frac{\gamma - \nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)\right) \mathrm{Q}^2\left(\frac{\gamma - \nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)$$
$$= 4\left(1 - \mathrm{Q}\left(\frac{\gamma - \nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)\right) \mathrm{Q}\left(\frac{\gamma - \nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right). \quad (102)$$

The third and fourth moment in normalized form are given by

$$\bar{\mu}_3(\theta) = \sum_{\mathcal{Z}} \left(\frac{z - \mu_1(\theta)}{\sqrt{\mu_2(\theta)}}\right)^3 p(z;\theta)$$
$$= \frac{1 - 2\,\mathrm{Q}\left(\frac{\gamma-\nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)}{\sqrt{\left(1 - \mathrm{Q}\left(\frac{\gamma-\nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)\right) \mathrm{Q}\left(\frac{\gamma-\nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)}} \quad (103)$$

and

$$\bar{\mu}_4(\theta) = \sum_{\mathcal{Z}} \left(\frac{z - \mu_1(\theta)}{\sqrt{\mu_2(\theta)}}\right)^4 p(z;\theta)$$
$$= \frac{1}{\left(1 - \mathrm{Q}\left(\frac{\gamma-\nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)\right) \mathrm{Q}\left(\frac{\gamma-\nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)} - 3. \quad (104)$$

With the derivatives

$$\frac{\partial \mu_1(\theta)}{\partial \theta} = \frac{2\mathrm{e}^{-\frac{(\gamma-\nu_1(\theta))^2}{2\nu_2(\theta)}} \left(\frac{\partial \nu_1(\theta)}{\partial \theta}\sqrt{\nu_2(\theta)} + \frac{(\gamma-\nu_1(\theta))}{2\sqrt{\nu_2(\theta)}}\frac{\partial \nu_2(\theta)}{\partial \theta}\right)}{\sqrt{2\pi}\nu_2(\theta)} \quad (105)$$

and

$$\frac{\partial \mu_2(\theta)}{\partial \theta} = \frac{4\mathrm{e}^{-\frac{(\gamma-\nu_1(\theta))^2}{2\nu_2(\theta)}} \left(\frac{\partial \nu_1(\theta)}{\partial \theta}\sqrt{\nu_2(\theta)} + \frac{(\gamma-\nu_1(\theta))}{2\sqrt{\nu_2(\theta)}}\frac{\partial \nu_2(\theta)}{\partial \theta}\right)}{\sqrt{2\pi}\nu_2(\theta)} \cdot$$
$$\cdot \left(1 - 2\,\mathrm{Q}\left(\frac{\gamma - \nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)\right), \quad (106)$$

we verify that

$$\frac{\partial \mu_1(\theta)}{\partial \theta}\sqrt{\mu_2(\theta)}\bar{\mu}_3(\theta) = \left(1 - 2\,\mathrm{Q}\left(\frac{\gamma - \nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)\right) \cdot$$
$$\cdot \frac{4\mathrm{e}^{-\frac{(\gamma-\nu_1(\theta))^2}{2\nu_2(\theta)}} \left(\frac{\partial \nu_1(\theta)}{\partial \theta}\sqrt{\nu_2(\theta)} + \frac{(\gamma-\nu_1(\theta))}{2\sqrt{\nu_2(\theta)}}\frac{\partial \nu_2(\theta)}{\partial \theta}\right)}{\sqrt{2\pi}\nu_2(\theta)}$$
$$= \frac{\partial \mu_2(\theta)}{\partial \theta}. \quad (107)$$

Therefore, the information bound is given by

$$S(\theta) = \frac{1}{\mu_2(\theta)}\left(\frac{\partial \mu_1(\theta)}{\partial \theta}\right)^2$$
$$= \frac{\mathrm{e}^{-\frac{(\gamma-\nu_1(\theta))^2}{\nu_2(\theta)}} \left(\frac{\partial \nu_1(\theta)}{\partial \theta}\sqrt{\nu_2(\theta)} + \frac{(\gamma-\nu_1(\theta))}{2\sqrt{\nu_2(\theta)}}\frac{\partial \nu_2(\theta)}{\partial \theta}\right)^2}{2\pi\nu_2^2(\theta)\,\mathrm{Q}\left(\frac{\gamma-\nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)\left(1 - \mathrm{Q}\left(\frac{\gamma-\nu_1(\theta)}{\sqrt{\nu_2(\theta)}}\right)\right)}. \quad (108)$$

Comparing this with the expression (100) for the exact information measure $F(\theta)$, it can be concluded that also for a generic hard-limited Gaussian distribution the information bound $S(\theta)$ is a pessimistic approximation for the Fisher information $F(\theta)$ with extraordinary quality.

VI. Applications

Finally, we want to outline possible applications of the presented approach and the opportunities provided by an information bound like $S(\theta)$. To this end, we present three problems for which $S(\theta)$ provides interesting and useful insights. The discussed problems cover theoretic as well as practical aspects in statistical signal processing.

A. Worst-Case Noise and Minimum Fisher Information

An important question in signal processing is to specify the worst-case noise distribution under the considered system model [25]. A common assumption in the field is that noise affects technical receive systems in an additive way. Therefore a model of high practical relevance is

$$Z = x(\theta) + W, \quad (109)$$

where $x(\theta)$ is a deterministic pilot signal modulated by the unknown parameter θ (for example attenuation, time-delay, frequency-offset, etc.) and W is additive independent random noise. Without loss of generality it can be assumed that the noise is zero mean, i.e.,

$$\mathrm{E}\left[W\right] = 0. \quad (110)$$

If in addition the noise has the property

$$\mathrm{E}\left[W^2\right] = \nu, \quad (111)$$

i.e., the second central moment of Z is constant, it is well-understood, that assuming the noise component W to follow the Gaussian probability density function

$$p(w) = \frac{1}{\sqrt{2\pi\nu}}\mathrm{e}^{-\frac{w^2}{2\nu}}, \quad (112)$$

leads to minimum Fisher information $F(\theta)$ [26] [11]. Therefore, under an estimation theoretic perspective, Gaussian noise is the worst-case assumption in an additive system like (109) with constant second output moment [21]. The presented bounding approach $S(\theta)$ allows slightly stronger statements.

If for any system $p(z;\theta)$ (including non-additive systems) the output Z exhibits the characteristic

$$\begin{aligned}\mu_1(\theta) &= \mathrm{E}\left[Z\right]\\ &= x(\theta), \end{aligned} \tag{113}$$
$$\begin{aligned}\mu_2(\theta) &= \mathrm{E}\left[(Z-\mu_1(\theta))^2\right]\\ &= \nu, \end{aligned} \tag{114}$$

the presented result shows that $F(\theta)$ can not violate

$$F(\theta) \geq \frac{1}{\mu_2(\theta)} \frac{\left(\frac{\partial \mu_1(\theta)}{\partial\theta}\right)^2}{1-\frac{\bar{\mu}_3^2(\theta)}{(\bar{\mu}_4(\theta)-1)}}. \tag{115}$$

This lower bound is minimized by a symmetric distribution, i.e., $\bar{\mu}_3(\theta)=0$. The resulting expression

$$F(\theta) \geq \frac{1}{\mu_2(\theta)}\left(\frac{\partial \mu_1(\theta)}{\partial\theta}\right)^2, \tag{116}$$

reaches equality under an additive Gaussian system model

$$p(z;\theta) = \frac{1}{\sqrt{2\pi\nu}}\mathrm{e}^{-\frac{(z-x(\theta))^2}{2\nu}}, \tag{117}$$

such that the worst-case model assumption with respect to Fisher information under the considered restrictions (113) and (114) is in general additive and Gaussian. In the more general setting, where also the second output moment exhibits a dependency on the system parameter θ,

$$\begin{aligned}\mu_1(\theta) &= \mathrm{E}\left[Z\right]\\ &= x(\theta), \end{aligned} \tag{118}$$
$$\begin{aligned}\mu_2(\theta) &= \mathrm{E}\left[(Z-\mu_1(\theta))^2\right]\\ &= \nu(\theta) \end{aligned} \tag{119}$$

and additionally the output distribution is symmetric, i.e.,

$$\bar{\mu}_3(\theta) = 0, \tag{120}$$

the presented result allows to conclude, that the Fisher information is in general bounded from below by

$$F(\theta) \geq \frac{1}{\nu(\theta)}\left(\frac{\partial x(\theta)}{\partial\theta}\right)^2 + \frac{1}{\nu^2(\theta)(\bar{\mu}_4(\theta)-1)}\left(\frac{\partial \nu(\theta)}{\partial\theta}\right)^2. \tag{121}$$

As the system model

$$p(z;\theta) = \frac{1}{\sqrt{2\pi\nu(\theta)}}\mathrm{e}^{-\frac{(z-x(\theta))^2}{2\nu(\theta)}} \tag{122}$$

exhibits the inference capability

$$F(\theta) = \frac{1}{\nu(\theta)}\left(\frac{\partial x(\theta)}{\partial\theta}\right)^2 + \frac{1}{2\nu^2(\theta)}\left(\frac{\partial \nu(\theta)}{\partial\theta}\right)^2, \tag{123}$$

it can be concluded together with (27) that for all cases where

$$1 \leq \bar{\mu}_4(\theta) \leq 3, \tag{124}$$

the worst-case system model $p(z;\theta)$ with respect to parameter estimation is the parametric Gaussian one (122).

B. Information Loss - Squaring Device

Another interesting problem in statistical signal processing is to characterize the estimation theoretic quality of non-linear receive and measurement systems. The Fisher information measure $F(\theta)$ is a rigorous tool which allows to draw precise conclusions. However, depending on the nature of the non-linearity, the exact calculation of the information measure $F(\theta)$ can become complicated. As an example for such a scenario consider the problem of analyzing the intrinsic capability of a system with a squaring sensor output

$$Z = Y^2, \tag{125}$$

to infer the mean θ of a Gaussian input

$$p(y;\theta) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(y-\theta)^2}{2}} \tag{126}$$

with unit variance. In such a case the system output Z follows a non-central chi-square distribution parameterized by θ. As the analytical description of the associated probability density function $p(z;\theta)$ includes a Bessel function, the characterization of the Fisher information $F(\theta)$ in compact analytical form is non-trivial. We short-cut the derivation by using the presented approximation $S(\theta)$ instead of $F(\theta)$. The first two output moments are found to be given by

$$\begin{aligned}\mathrm{E}\left[Z\right] &= \mathrm{E}\left[\theta^2+2\theta W+W^2\right]\\ &= \theta^2+1\\ &= \mu_1(\theta) \end{aligned} \tag{127}$$
$$\begin{aligned}\mathrm{E}\left[(Z-\mu_1(\theta))^2\right] &= \mathrm{E}\left[(\theta^2+2\theta W+W^2-\theta^2-1)^2\right]\\ &= 2(2\theta^2+1)\\ &= \mu_2(\theta), \end{aligned} \tag{128}$$

where we have introduced the auxiliary random variable

$$W = Y-\theta. \tag{129}$$

The third output moment is

$$\begin{aligned}\mathrm{E}\left[(Z-\mu_1(\theta))^3\right] &= \mathrm{E}\left[(\theta^2+2\theta W+W^2-\theta^2-1)^3\right]\\ &= 8(3\theta^2+1)\\ &= \mu_3(\theta), \end{aligned} \tag{130}$$

while the fourth moment is

$$\begin{aligned}\mathrm{E}\left[(Z-\mu_1(\theta))^4\right] &= \mathrm{E}\left[(\theta^2+2\theta W+W^2-\theta^2-1)^4\right]\\ &= 12\big((2\theta^2+1)^2+4(4\theta^2+1)\big)\\ &= \mu_4(\theta). \end{aligned} \tag{131}$$

The normalized versions of the third and fourth moment are

$$\begin{aligned}\bar{\mu}_3(\theta) &= \mu_3(\theta)\mu_2^{-\frac{3}{2}}(\theta)\\ &= \frac{8(3\theta^2+1)}{2\sqrt{2}(2\theta^2+1)^{\frac{3}{2}}}\\ &= \frac{2\sqrt{2}(3\theta^2+1)}{(2\theta^2+1)^{\frac{3}{2}}} \end{aligned} \tag{132}$$

and

$$\begin{aligned}\bar{\mu}_4(\theta) &= \mu_4(\theta)\mu_2^{-2}(\theta)\\ &= \frac{12\big((2\theta^2+1)^2+4(4\theta^2+1)\big)}{4(2\theta^2+1)^2}\\ &= \frac{12(4\theta^2+1)}{(2\theta^2+1)^2}+3. \end{aligned} \tag{133}$$

With the derivatives

$$\begin{aligned}\frac{\partial\mu_1(\theta)}{\partial\theta} &= 2\theta\\ \frac{\partial\mu_2(\theta)}{\partial\theta} &= 8\theta,\end{aligned} \tag{134}$$

we obtain

$$\begin{aligned}\beta^\star(\theta) &= \frac{\frac{\partial\mu_1(\theta)}{\partial\theta}\sqrt{\mu_2(\theta)}\bar{\mu}_3(\theta)-\frac{\partial\mu_2(\theta)}{\partial\theta}}{\frac{\partial\mu_2(\theta)}{\partial\theta}\bar{\mu}_3(\theta)-\frac{\partial\mu_1(\theta)}{\partial\theta}\sqrt{\mu_2(\theta)}(\bar{\mu}_4(\theta)-1)}\\ &= -\frac{\theta^2\sqrt{2}\sqrt{(2\theta^2+1)}}{(4\theta^4+16\theta^2+3)}\end{aligned} \tag{135}$$

and the approximation is finally given by

$$\begin{aligned}S(\theta) &= \frac{1}{\mu_2(\theta)}\frac{\left(\frac{\partial\mu_1(\theta)}{\partial\theta}+\frac{\beta^\star(\theta)}{\sqrt{\mu_2(\theta)}}\frac{\partial\mu_2(\theta)}{\partial\theta}\right)^2}{1+2\beta^\star(\theta)\bar{\mu}_3(\theta)+\beta^{\star 2}(\theta)(\bar{\mu}_4(\theta)-1)}\\ &= \frac{2\theta^2\left(4\theta^4+12\theta^2+3\right)^2}{\left(4\theta^4+12\theta^2+3\right)\left(8\theta^6+24\theta^4+18\theta^2+3\right)}\\ &= \frac{2\theta^2\left(4\theta^4+12\theta^2+3\right)}{\left(8\theta^6+24\theta^4+18\theta^2+3\right)}.\end{aligned} \tag{136}$$

Fig. 1 depicts the approximative information loss

$$\tilde{\chi}(\theta) = \frac{S_Z(\theta)}{F_Y(\theta)}, \tag{137}$$

when squaring the random input variable Y. As a comparison also the corresponding loss for a symmetric hard-limiter (91) with $\gamma = 0$ is visualized. It can be observed that for low values of θ the information about the sign (hard-limiting) of the system input Y conveys much more information about the input mean θ than the amplitude (squaring). For $\theta \geq 0.75$ the situation changes and the squaring receiver outperforms the hard-limiter when it comes to estimating the mean θ of the input Y from samples of the system output Z.

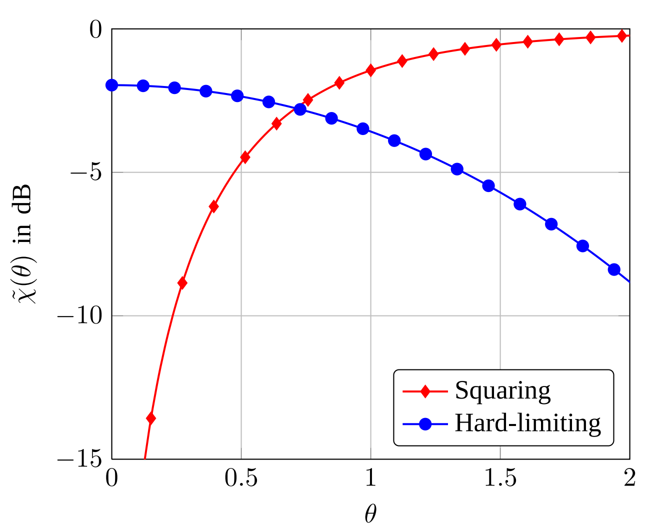

Fig. 1. Non-linear Systems - Performance Loss

C. Measuring Inference Capability - Soft-Limiter

A situation that can be encountered in practice is that the analytical characterization of the system model $p(z;\theta)$ or its moments is difficult. If the appropriate parametric system model $p(z;\theta)$ is unknown, the direct consultation of an analytical tool like the Fisher information measure $F(\theta)$ becomes impossible. However, in such a situation the presented approach of the information bound $S(\theta)$ allows to numerically approximate the Fisher information measure $F(\theta)$ at low-complexity. To this end, the moments of the system output Z are measured in a calibrated setup, where the parameter θ can be controlled, or determined by Monte-Carlo simulations. We demonstrate this validation technique by using a soft-limiter model, i.e., the system input Y is transformed by

$$\begin{aligned}Z &= \sqrt{\frac{2}{\pi\zeta^2}}\int_0^Y \mathrm{e}^{-\frac{u^2}{2\zeta^2}}\mathrm{d}u\\ &= \mathrm{erf}\left(\frac{Y}{\sqrt{2\zeta^2}}\right),\end{aligned} \tag{138}$$

where $\zeta \in \mathbb{R}$ is a constant model parameter and

$$\mathrm{erf}\,(x) = \frac{2}{\sqrt{\pi}}\int_0^x \mathrm{e}^{-t^2}\mathrm{d}t \tag{139}$$

is the error function. This non-linear model can for example be used in order to characterize saturation effects in analog system components like low-noise amplifiers. In Fig. 2 the input-to-output mapping of the model (138) is depicted for different setups ζ. As input we consider a Gaussian distribution with unit variance like in (126). The output moments $\mu_1(\theta), \mu_2(\theta), \bar{\mu}_3(\theta), \bar{\mu}_4(\theta)$ are measured by simulating the

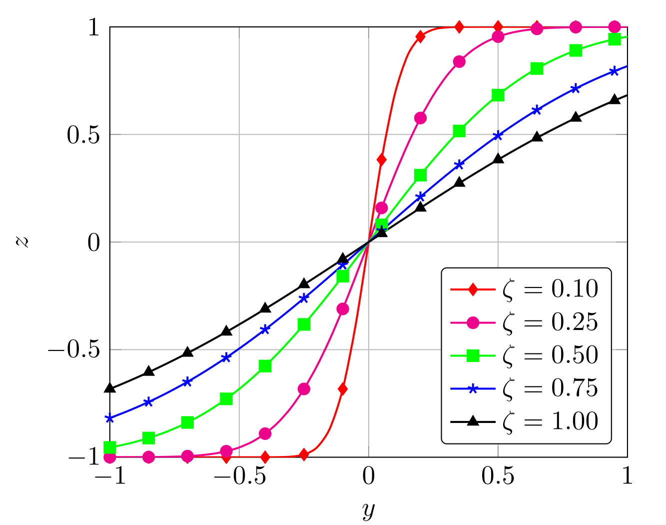

Fig. 2. Soft Limiter Model - Input-to-Output

non-linear system output Z with 10^9 independent realizations for each considered value of the input mean θ. The result is shown in Fig. 3. After numerically approximating the required

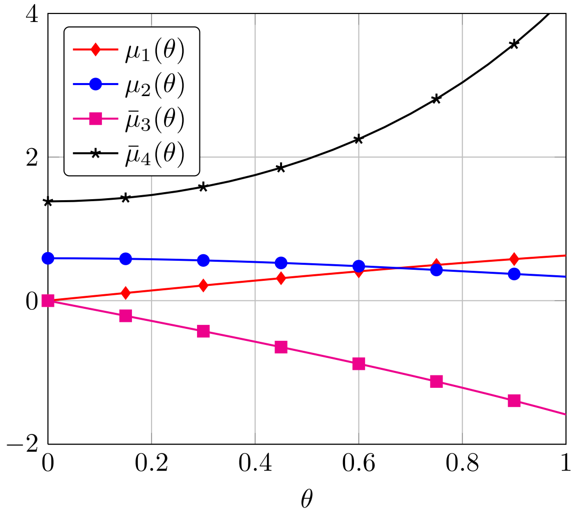

Fig. 3. Soft-Limiter Model - Measured Moments ($\zeta = 0.5$)

derivatives $\frac{\partial \mu_1(\theta)}{\partial \theta}, \frac{\partial \mu_2(\theta)}{\partial \theta}$, which are depicted in Fig. 4, the approximation $S(\theta)$ is calculated. In Fig. 5 the measured

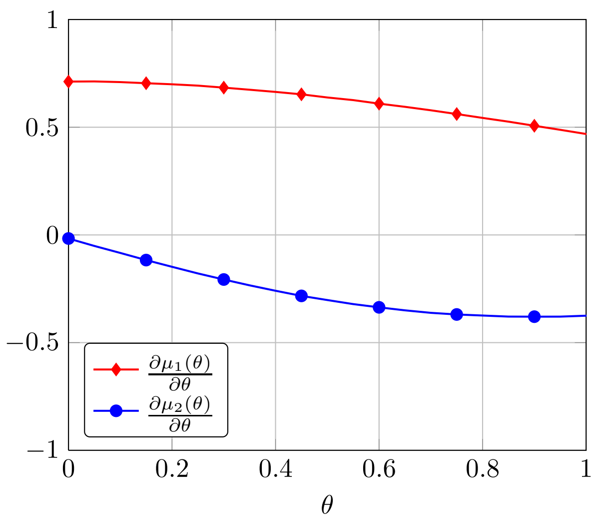

Fig. 4. Soft-Limiter Model - Measured Derivatives ($\zeta = 0.5$)

information loss $\tilde{\chi}(\theta)$ of the soft-limiter model is shown, where the dotted line indicates the exact information loss $\chi(\theta)$ with a hard-limiter (91) which is equivalent to a soft-limiter with $\zeta \to 0$.

VII. Conclusion

We have established a strong and generic lower bound for the Fisher information measure. By various examples we have shown that the derived expression has the potential to provide a good approximation in a broad number of cases. This makes the presented information bound a versatile mathematical tool for a variety of problems encountered in the design and optimization of signal processing systems. Further, the pessimistic

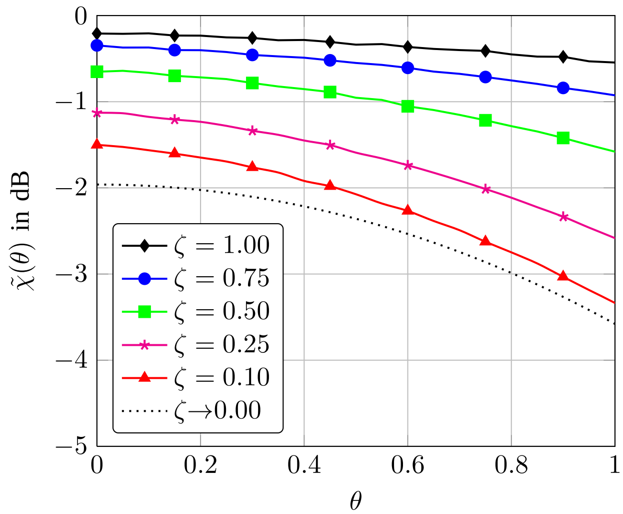

Fig. 5. Soft-Limiter Model - Information Loss

nature of the attained alternative information measure allows to strengthen insights on worst-case noise and to generalize classical results on Gaussian system models which exhibit minimum Fisher information. Finally, we have outlined how to use the presented information bound in order to benchmark physical measurement systems with output statistics of unknown analytical form.

References

[1] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philosoph. Trans. Roy. Soc. London*, vol. 222, pp. 309–368, Jan. 1922.

[2] R. A. Fisher, "Theory of statistical estimation," *Proc. Cambridge Philosph. Soc.*, vol. 22, no.5, pp. 700–725, Jul. 1925.

[3] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddler River, NJ: Prentice Hall, 1993.

[4] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, no.3, pp. 81–91, 1945.

[5] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1946.

[6] A. Bhattacharyya , "On Some Analogues of the Amount of Information and Their Use in Statistical Estimation," *Sankhya: The Indian Journal of Statistics*, vol. 8, no. 1, pp. 1–14, 1946.

[7] E. W. Barankin, "Locally Best Unbiased Estimates," *The Annals of Mathematical Statistics*, vol. 22, no. 4, pp. 477–501, 1949.

[8] J. M. Hammersley, "On Estimating Restricted Parameters," *Journal of the Royal Statistical Society* vol. 12, no. 2, pp. 192–240,1950.

[9] D. G. Chapman, H. Robbins, "Minimum Variance Estimation Without Regularity Assumptions," *The Annals of Mathematical Statistics*, vol. 22, no. 4, pp. 581–586, 1951.

[10] H. Kobayashi, B. L. Mark, W. Turin, *Probability, Random Processes, and Statistical Analysis*. Cambridge University Press, 2011.

[11] D. E. Boekee, "A generalization of the Fisher information measure," Ph. D. dissertation, TU Delft, Delft, Netherlands, 1977.

[12] H. L. Van Trees *Detection, Estimation, and Modulation Theory, Part I*, New York: Wiley, 1968.

[13] J. Ziv, M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Transactions on Information Theory*, vol. 15, no. 3, pp. 386–391, 1969.

[14] B. Bobrovsky, M. Zakai, "A lower bound on the estimation error for Markov processes," *IEEE Transactions on Automatic Control*, vol. 20, no. 6, pp. 785-788, 1975.

[15] A. J. Weiss, E. Weinstein, "A lower bound on the mean-square error in random parameter estimation," *IEEE Transactions on Information Theory*, vol. 31, no. 5, pp. 680–682, 1985.

[16] H. L. Van Trees, K. L. Bell, "Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking," 1st ed. Piscataway, NJ: Wiley-IEEE Press, 2007.

[17] M. Stein, A. Mezghani, J. A. Nossek, ”A lower bound for the Fisher information measure,” *IEEE Signal Process. Letters*, vol. 21, no. 7, pp. 796–799, July 2014.

[18] E. Uhrmann-Klingen, ”Minimal Fisher information distributions with compact-supports,” *Sankhyā: Indian J. Stat.*, vol. 57, no. 3, pp. 360-374, Oct. 1995.

[19] V. Živojnović, ”Minimum Fisher information of moment-constrained distributions with application to robust blind identification,” *Elsevier Signal Process.*, vol. 65, no. 2, pp. 297 – 313, Oct. 1998.

[20] J. F. Bercher, C. Vignat, ”On minimum Fisher information distributions with restricted support and fixed variance,” *Inf. Sci.*, vol. 179, no. 22, Nov. 2009.

[21] P. Stoica, P. Babu, ”The Gaussian data assumption leads to the largest Cramèr-Rao bound,” *IEEE Signal Process. Mag.*, vol. 28, pp. 132–133, May 2011.

[22] K. Pearson, Mathematical Contributions to the Theory of Evolution. XIX. Second Supplement to a Memoir on Skew Variation, *Phil.Trans. Roy. Soc. (A)*, vol. 216, pp. 429–457, 1916.

[23] J. E. Wilkins, ”A Note on Skewness and Kurtosis,” *The Annals of Mathematical Statistics*, vol. 15, no. 3, pp. 333-335, 1944.

[24] J. H. Van Vleck, D. Middleton, ”The spectrum of clipped noise,” *Proc. IEEE*, vol. 54, no. 1, pp. 2–19, 1966.

[25] P. J. Huber, ”Robust statistics: a review,” *Ann. Math. Statist.*, vol. 43, no. 4, pp. 1041–1067, 1972.

[26] A. J. Stam, ”Some mathematical properties of quantities of information,” Ph.D. dissertation, TU Delft, Delft, Netherlands, 1959.